AI's math gains trigger a fight over what progress is for

AI News Digest

2026-09-12

AI's math gains trigger a fight over what progress is for

By AI News Digest • September 12, 2026

A Fields Medallists' declaration challenges benchmark-driven mathematics, while AI-science automation, enterprise data policy, safety proposals, and ChatGPT Sites show capability colliding with institutional limits.

The main signal

Mathematics is where AI output and understanding are visibly pulling apart

A declaration signed by 25 Fields Medallists says LLMs can now solve major outstanding problems across mathematics, but argues that turning those problems into AI benchmarks is detrimental to the discipline. The signatories say solving is only a proxy for conceptual understanding and insight; rushed AI-generated solutions can arrive without proper writeups, attribution, or the human work needed to integrate ideas into the mathematical canon. They still acknowledge that AI could enhance and accelerate genuine mathematical study, making this a dispute over purpose and institutions rather than a denial of capability. [1, 2]

The counterargument is that AI may also remove a longstanding bottleneck. Emad Mostaque argues that models can now write Lean proofs that are checked in hours rather than years, turning correctness into a portable certificate—but leaving assumptions, importance, and the sorting of deep from shallow results to human judgment. The practical question is therefore shifting from whether an agent can produce a proof to who controls verification, credit, and the choice of questions worth asking. [1]

Agents move into institutional workflows

Inherent is building an AI-science organization, not just a research agent

Inherent says it has raised $50 million from Index Ventures and Radical Ventures at a $225 million post-money valuation with an 11-person team. The company grew out of DeepMind's AI Scientist effort and says its goal is a horizontal intelligence layer for science, giving agents human-equivalent context and affordances while redesigning how scientists and agents work together. [3]

Its Faraday system uses a 27B Qwen 3.6 model that can call a frontier coding agent as a tool. On Inherent's Replica task space, it was trained on 242 of 300 figure-replication tasks and tested on 68 held-out AI-for-science tasks under a one-hour, one-seventh-H200 budget; Inherent reports that it outperformed the coding agent it used, Claude, and GLM 5.2, while prompt optimization improved GPT-5.5 Codex only slightly. [3]

The result is bounded: the task set favors well-known, highly cited papers, and the team says it still needs to handle non-replicable work without rewarding cheating or Goodharting. Inherent also says Faraday is not yet general or reliable enough for public release, but is already used internally as part of a company-level loop in which humans and agents recursively improve the organization. [3]

Anthropic's enterprise data reversal is more consequential than the Fable label

According to Ben Thompson's account of the release, Anthropic stressed that Fable 5.1 is the same model as Mythos 5.1 and positioned it for lengthy software projects, code reviews, experiment design, and work with dense diagrams and tables. The more consequential move is Anthropic's rollback of a controversial business-customer data-retention policy after customer feedback: it plans to replace it with Enterprise Frontier Safeguards, and Fable is to operate with zero data retention until that system is available in a few months. [4]

Thompson interprets the reversal as a market check. Enterprise buyers treat control of proprietary data as foundational, and he argues that aggressive competition from OpenAI helped make Anthropic's policy untenable. For adoption, the durable differentiator is not just model capability but whether customers can keep control of the information used to generate it. [4]

Governance follows the capability curve

Safety politics split between a future ban and present-day controls

Alex Sobel says he and 71 colleagues are asking the UK government to work with them on a bill prohibiting the development of superintelligent AI and on

an international agreement, while preserving the country's strategic AI ambitions. The source presents this as a call for government cooperation on proposed legislation, not as an enacted rule. [5]

At the operational end of the debate, former Anthropic safety-team member Joe Benton says he has joined METR Evals to conduct independent evaluations. He calls for disclosure of progress toward recursive self-improvement, safety incidents and near-misses, minimum safety standards, and independent guarantees that companies meet them. Gary Marcus supplies the opposing emphasis: the immediate problem, he argues, is reckless and unreliable AI already deployed online, not AGI. [6, 7]

Product adoption

ChatGPT Sites passes five million creations

OpenAI says ChatGPT Sites passed five million user-created sites roughly three months after launch. The product now supports collaborative editing, private sharing, faster prompt-to-deployment, database inspection, and custom domains—evidence that ChatGPT is being used as an app-building and hosting surface as well as a conversational interface. [8]

Sources

1. X article by @EMostaque
2. Declaration — Math and AI
3. Why Scientific Taste Must Be Learned Through Practice — Edward Hughes
4. Fable 5.1 and Anthropic's Data Retention Pivot | Sharp Tech with Ben Thompson
5. Alex Sobel MP for Leeds Central and Headingley (@alexsobel) on X
6. X post by @JoeJBenton
7. X post by @GaryMarcus
8. X post by @ChatGPT